



JACOBS
UNIVERSITY

Prof. Dr. Michael Kohlhase, Jacobs Univ., 28759 Bremen, Germany

Charles University Prague
Faculty of Mathematics and Physics
The Dean
Ke Karlovu 3
121 16 Praha 2
Czech Republic

Prof. Dr. Michael Kohlhase
Professor for Computer Science

Campus Ring 1
D-28759 Bremen, Germany

Tel: +49 421 200-3140

Fax: +49 421 200-493140

m.kohlhase@jacobs-
university.de

Bremen, 2. Juli 2013

Evaluation of the Dissertation “*Mathematical Search Engine*” by Jozef Mižsutka

Topic

The dissertation contributes to the field of mathematical information retrieval (MIR). MIR has often been touted as one of the “killer applications” of computer support in STEM (Science, Technology, Engineering, and Mathematics). Indeed, it has been estimated that the yearly production of published research articles in mathematics exceeds half a million pages a year, the amount of internal technical documents in company archives certainly exceeds this number by far, accessing them will become a determining factor. Scientifically, MIR is an interesting and innovative new field at the intersection of information retrieval, knowledge representation, and digitization efforts. I expect MIR to give valuable stimuli for research to all of them.

Contribution

The work presented in this thesis contributes to MIR in two ways: Jozef Mižsutka has built (and optimized) a concrete search engine EgoMath, and he has initiated scientific evaluation for MIR systems. There are currently three approaches to MIR: similarity search by verbalizing formulae, feature-based search, and unification search. The EgoMath search engine is based on the first approach. The author has started development on this system in the second wave of MIR systems (around 2006). Like other systems in its class, it treats mathematical formulae as “special words” by verbalizing (their important aspects) and indexes them in a conventional bag-of-words search engine (here EgoThor from Prague University). The EgoMath system invests heavily into an input subsystem that can process formulae in various input formulae. The input system is essentially independent of EgoMath, and could thus be used by other systems. I attribute the fact that this useful system has not been taken up in other projects to the lack of early publication of its availability.

Evaluation

Form: The thesis begins with a thorough recap of the state of the art and history of MIR systems (Chapters 2. and 3.); this is the most complete and thorough such survey I have seen to date and

constitutes an independent scientific achievement. The thesis is well-structured, and substantiates all claims either from the literature or by carefully documented performance experiments. But I must say that the role of the multiple appendixes (especially appendix J with an early paper on EgoMath) did not quite become clear to me. There was some duplication of argumentation, but not to a degree that becomes problematic.

Research Contribution: Even though the ideas, architecture, and algorithms of the EgoMath presented in Chapter 4. are rather standard now, at the time they were innovative and state of the art when they were developed and have been updated since then.

The really novel part of the thesis is the comparative evaluation of MIR systems in Chapter 5. To the best of my knowledge this is the first systematic comparison of two MIR systems (I consider the MIR happening I organized at CICM 2012 to be unsystematic) and Jozef Mižsutka has re-implemented a feature-based MIR system for the comparison. Jozef Mižsutka presents detailed experimental performance measurements. I would have liked to have more concrete conclusions drawn from this that would guide further research. The only hint at this we get is that the results have been used to improve the EgoMath search engine.

Summary

The thesis presented by Jozef Mižsutka presents a solid piece of scientific work and shows the ability of Jozef Mižsutka to think creatively and systematically on innovative subjects. The work presented in the thesis has room for improvement in details, but meets the requirements for a Ph.D. thesis.

I recommend the thesis “Mathematical Search Engine” by Jozef Mižsutka for acceptance.

Michael Kohlhase

Prof. Dr. Michael Kohlhase